ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Linear Algebra and its Applications 419 (2006) 53–77

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

# Perturbation of the SVD in the presence of small singular values

Michael Stewart

*Department of Mathematics and Statistics, Georgia State University, 30 Pryor St., Atlanta, GA 30303, United States*

Received 12 August 2004; accepted 3 April 2006

Available online 30 June 2006

Submitted by V. Mehrmann

**Abstract**

This paper gives SVD perturbation bounds and expansions that are of use when an $m \times n$, $m \geqslant n$ matrix $A$ has small singular values. The first part of the paper gives subspace bounds that are closely related to those of Wedin but are stated so as to isolate the effect of any small singular values to the left singular subspace. In the second part first and second order approximations are given for perturbed singular values. The subspace bounds are used to show that all approximations retain accuracy when applied to small singular values. The paper concludes by deriving a subspace bound for multiplicative perturbations and using that bound to give a simple approximation to a singular value perturbed by a multiplicative perturbation.



*Keywords:* Perturbation theory; Singular value; Singular subspace

## 1. Background

Suppose that the $m \times n$ matrix $A$ with $m \geqslant n$ has singular value decomposition

$$A = \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix} \tag{1}$$

and that a perturbed version of $A$, $\widetilde{A}$, has singular value decomposition

*E-mail address:* mastewart@gsu.edu

0024-3795/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2006.04.013

$$\widetilde{A} = \begin{bmatrix} \widetilde{U}_1 & \widetilde{U}_2 & \widetilde{U}_3 \end{bmatrix} \begin{bmatrix} \widetilde{\Sigma}_1 & 0 \\ 0 & \widetilde{\Sigma}_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{V}_1^H \\ \widetilde{V}_2^H \end{bmatrix}. \tag{2}$$

The blocks $\Sigma_1$ and $\Sigma_2$ are $m_1 \times m_1$ and $m_2 \times m_2$ respectively and the perturbed decomposition is partitioned in the same way. We assume that $\Sigma_1$, $\widetilde{\Sigma}_1$, $\Sigma_2$ and $\widetilde{\Sigma}_2$ are diagonal with positive diagonal elements. We will assume that $A$ is a complex matrix.

With each of the blocks of $U$ and $V$ we associate the singular subspaces

$$\mathscr{U}_k = \mathscr{R}(U_k), \quad \mathscr{V}_k = \mathscr{R}(V_k), \quad \widetilde{\mathscr{U}}_k = \mathscr{R}(\widetilde{U}_k) \quad \text{and} \quad \widetilde{\mathscr{V}}_k = \mathscr{R}(\widetilde{V}_k).$$

The goal of this paper is to describe the difference between the sets of perturbed and unperturbed singular values $\sigma(\widetilde{\Sigma}_1)$ and $\sigma(\Sigma_1)$, the difference between the left singular subspaces $\mathscr{U}_1$ and $\widetilde{\mathscr{U}}_1$ and the difference between the right singular subspaces $\mathscr{V}_1$ and $\widetilde{\mathscr{V}}_1$. Closely related subspace perturbation bounds can be found in [9]. Bounds and expansions for singular values and subspaces are surveyed in [6]. The motivation for presenting a new analysis is to derive theorems that are more accurate when $\Sigma_1$ has small singular values.

Small singular values complicate perturbation theory for both singular values and singular subspaces. For subspaces this can be seen in the following theorem from [9].

**Theorem 1** (Wedin). *Suppose that* $\delta$, $\alpha$ *and* $\beta$ *satisfying* $0 < \delta \leqslant \alpha \leqslant \beta$ *are such that* $\sigma(\widetilde{\Sigma}_1)$ *lies in* $[\alpha, \beta]$ *while* $\sigma(\Sigma_2)$ *lies outside of* $(\alpha - \delta, \beta + \delta)$. *Then*

$$\|\widetilde{U}_1^H \begin{bmatrix} U_2 & U_3 \end{bmatrix}\| \leqslant k \frac{\max(\|R\|, \|S\|)}{\delta}$$

*and*

$$\|\widetilde{V}_1^H V_2\| \leqslant k \frac{\max(\|R\|, \|S\|)}{\delta},$$

*where*

$$k = \begin{cases} \sqrt{2} & \text{when } \|\cdot\| \text{ is } \|\cdot\|_2 \text{ or } \|\cdot\|_F. \\ 2 & \text{when } \|\cdot\| \text{ is an arbitrary unitarily invariant norm.} \end{cases}$$

The size of the perturbation of $A$ is measured by the size of the residuals

$$R = A\widetilde{V}_1 - \widetilde{U}_1\widetilde{\Sigma}_1 \quad \text{and} \quad S = A^H\widetilde{U}_1 - \widetilde{V}_1\widetilde{\Sigma}_1. \tag{3}$$

The quantities bounded in Theorem 1 are the matrices of sines of the canonical angles between the perturbed and unperturbed subspaces [1].

The theorem assumes that $\sigma_j(\widetilde{\Sigma}_1) \geqslant \delta$. If $\widetilde{\Sigma}_1$ has small singular values then $\delta$ must be correspondingly small and the bounds suggest that the singular subspaces are sensitive. When a bound on the change in $\mathscr{U}_1$ is desired some assumption of this kind is expected: if any of the singular values of $\widetilde{\Sigma}_1$ are close to zero then the corresponding singular vectors are not well distinguished from the left null space $\mathscr{U}_3$.

In Section 2 we will avoid this difficulty by separating the bound on $U_3^H\widetilde{U}_1$ from those on $U_2^H\widetilde{U}_1$ and $V_2^H\widetilde{V}_1$, thereby eliminating the assumption $\sigma_j(\widetilde{\Sigma}_1) > \delta$ from all bounds on $U_2^H\widetilde{U}_1$ and $V_2^H\widetilde{V}_1$. In all other respects the subspace results are standard. The technique for deriving the bounds is that of [1] and the resulting theorems each have a counterpart from [9].

Perturbation expansions involving small singular values also pose difficulties. Suppose for the moment that $\Sigma_1 = \sigma_1$ and $\widetilde{\Sigma}_1 = \tilde{\sigma}_1$ are $1 \times 1$. If $\widetilde{A} = A + E$, $\sigma_1 \neq 0$ and $\sigma_1$ is distinct from the singular values of $\Sigma_2$ then it is known, [5], that for a real matrix and perturbation:

$$\tilde{\sigma}_1 = \sigma_1 + u_1^{\mathrm{T}} E v_1 + \mathrm{O}(\|E\|^2), \tag{4}$$

where $u_1$ and $v_1$ are the left and right singular vectors associated with $\sigma_1$. The essential assumption $\sigma_1 \neq 0$ ensures that $\tilde{\sigma}_1$ is a differentiable function of the elements of $E$. If $\sigma_1$ is small but nonzero the first order approximation is inaccurate. A natural way to attempt to restore accuracy is to keep track of second order terms [8,7]. It is also possible to derive an accurate second order expansion for $\sigma_1^2$ [5].

Unfortunately higher order expansions can suffer from the same problem. Consider the expansion:

$$\begin{aligned}\tilde{\sigma}_1 = {} & \sigma_1 + u_1^{\mathrm{T}} E v_1 + \frac{1}{2} \begin{bmatrix} u_1^{\mathrm{T}} & v_1^{\mathrm{T}} \end{bmatrix} \\ & \times \left\| \begin{bmatrix} \sigma_1 E V_2 (\sigma_1^2 I - \Sigma_2^{\mathrm{T}} \Sigma_2)^{-1} V_2^{\mathrm{T}} E^{\mathrm{T}} & E V_2 (\sigma_1^2 I - \Sigma_2^{\mathrm{T}} \Sigma_2)^{-1} \Sigma_2^{\mathrm{T}} U_2^{\mathrm{T}} E \\ E^{\mathrm{T}} U_2 \Sigma_2 (\sigma_1^2 I - \Sigma_2^{\mathrm{T}} \Sigma_2)^{-1} V_2^{\mathrm{T}} E^{\mathrm{T}} & \sigma_1 E^{\mathrm{T}} U_2 (\sigma_1^2 I - \Sigma_2 \Sigma_2^{\mathrm{T}})^{-1} U_2^{\mathrm{T}} E \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \right\|_{\mathrm{F}} \\ & + \mathrm{O}(\|E\|^3)\end{aligned} \tag{5}$$

from [7] where it is assumed that $\sigma_1 > 0$ and that the singular value $\sigma_1$ is distinct from the singular values in $\Sigma_2$. We also consider the expansion:

$$\tilde{\sigma}_1^2 = (\sigma_1 + u_1^{\mathrm{T}} E v_1)^2 + \|U_2^{\mathrm{T}} E v_1\|^2 + \|U_3^{\mathrm{T}} E v_1\|^2 + h^{\mathrm{T}} (\sigma_1^2 I - \Sigma_2^2)^{-1} h + \mathrm{O}(\|E\|^3), \tag{6}$$

from [5] where

$$h = \sigma_1 V_2^{\mathrm{T}} E^{\mathrm{T}} u_1 + \Sigma_2 U_2^{\mathrm{T}} E v_1.$$

This latter expansion does not require $\sigma_1 > 0$.

We illustrate potential problems with a simple example.

**Example 1**

$$A = \Sigma = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right] = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & 1 \\ 0 & 0 \end{array}\right] \quad \text{and} \quad \widetilde{A} = \Sigma + E = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & 1 \\ \epsilon & 0 \end{array}\right] \tag{7}$$

so that

$$\Sigma_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \epsilon & 0 \end{bmatrix}$$

with $U = I_{3\times 3}$ and $V = I_{2\times 2}$. Note that in order to apply a theorem from [7] using the notation in which it was originally presented we have temporarily changed the partitioning of $A$ so that $\Sigma_2$ is no longer square. Clearly $A$ has a singular value $\sigma_1$ and $\widetilde{A}$ has a singular value $\tilde{\sigma}_1 = \sqrt{\sigma_1^2 + \epsilon^2}$.

Applied to (7) the expansion (5) simplifies to

$$\tilde{\sigma}_1 = \sigma_1 + \frac{\epsilon^2}{2\sigma_1} + \mathrm{O}(\epsilon^3). \tag{8}$$

This approximation to the perturbed singular value is satisfactory so long as $\sigma_1$ is not comparable to or smaller than $\epsilon$. If $\sigma_1 = 10^{-10}$ and $\epsilon = 10^{-10}$ then (8) becomes

$$\tilde{\sigma}_1 = 10^{-10} + \frac{10^{-20}}{2 \times 10^{-10}} + \mathrm{O}(\epsilon^3) \approx 1.5 \times 10^{-10}.$$

The true singular value is

$$\tilde{\sigma}_1 = \sqrt{(10^{-10})^2 + (10^{-10})^2} \approx 1.4 \times 10^{-10}.$$

The simple first order expansion (4) gives $\tilde{\sigma}_1 \approx 10^{-10}$. If $\sigma_1 = 10^{-20}$ and $\epsilon = 10^{-10}$ then the results are worse. The true singular value is $\tilde{\sigma}_1 \approx 10^{-10}$ but (8) gives $\tilde{\sigma}_1 \approx 0.5$ and (4) gives $\tilde{\sigma}_1 \approx 10^{-20}$.

Applying (6) to (7) and simplifying yields

$$\tilde{\sigma}_1^2 = \sigma_1^2 + \epsilon^2 + \mathrm{O}(\epsilon^3),$$

which is the exact perturbed singular value.

The neglected terms in (6) really are $\mathrm{O}(\|E\|^3)$, even when $\sigma_1$ is small. However it is worth noting that an $\mathrm{O}(\|E\|^3)$ error in approximating $\tilde{\sigma}^2$ does not necessarily translate into an $\mathrm{O}(\|E\|^3)$ error in approximating $\tilde{\sigma}_1$. Suppose that $\breve{\sigma}_1^2 > 0$ is an approximation $\tilde{\sigma}_1^2$ with $\mathrm{O}(\epsilon^3)$ error so that

$$\tilde{\sigma}_1^2 = \breve{\sigma}_1^2 + \epsilon^3.$$

Then

$$\tilde{\sigma}_1 = \breve{\sigma}_1\sqrt{1 + \frac{\epsilon^3}{\breve{\sigma}_1^2}} = \breve{\sigma}_1\left(1 + \frac{\epsilon^3}{2\breve{\sigma}_1^2} + \mathrm{O}\left(\frac{\epsilon^6}{\breve{\sigma}_1^4}\right)\right) = \breve{\sigma}_1 + \frac{\epsilon^3}{2\breve{\sigma}_1} + \breve{\sigma}_1\mathrm{O}\left(\frac{\epsilon^6}{\breve{\sigma}_1^4}\right).$$

An approximation to $\tilde{\sigma}_1^2$ with error $\epsilon^3$ corresponds to an approximation to $\tilde{\sigma}_1$ with error that is $\mathrm{O}(\epsilon^3/\breve{\sigma}_1)$. If $\breve{\sigma}_1 = \mathrm{O}(\epsilon)$ then the result is an approximation that is not in general any better than a first order approximation. This will be observed in a numerical example in §4.

In Section 3, we will derive a first order approximation to a set of singular values in $\widetilde{\Sigma}_1$ under the assumption that they are well separated from those in $\Sigma_2$. We do not assume that the singular values in $\widetilde{\Sigma}_1$ are nonzero. The theorem is applicable to a group of singular values and the neglected terms admit a strict upper bound. The bound on the error is based on the subspace bounds in Section 2. In particular we will show that there are unitary matrices $Q$ and $P$ such that

$$\begin{bmatrix} \widetilde{\Sigma}_1 \\ 0 \end{bmatrix} = Q \begin{bmatrix} \Sigma_1 + U_1^H E V_1 \\ U_3^H E V_1 \end{bmatrix} P + H$$

with

$$\|H\|_2 \leqslant \frac{6}{\delta^2}\|E\|_2^2\|A + E\|_2 + \frac{2\sqrt{2}}{\delta}\|E\|_2\|E\|_2,$$

where $\delta$ is a measure of the separation between the singular values of $\widetilde{\Sigma}_1$ and those of $\Sigma_2$. It follows from Mirsky's theorem that the singular values of

$$\begin{bmatrix} \Sigma_1 + U_1^H E V_1 \\ U_3^H E V_1 \end{bmatrix}$$

are a first order approximation to those of $\widetilde{\Sigma}_1$. If $\widetilde{\Sigma}_1 = \tilde{\sigma}_1$ is $1 \times 1$ we get

$$\tilde{\sigma}_1 = \left\| \begin{bmatrix} \sigma_1 + u_1^H E v_1 \\ U_3^H E v_1 \end{bmatrix} \right\|_2 + \mathrm{O}(\|E\|^2).$$

This retains first order accuracy regardless of the size of $\tilde{\sigma}_1$. In Section 4 we give an analogous second order approximation to an individual perturbed singular value. The approximation retains second order accuracy regardless of the size of $\tilde{\sigma}_1$. Finally, in Section 5 we briefly consider multiplicative perturbations and give a subspace bound and an approximation to a singular value of the perturbed matrix $\widetilde{A} = A(I + E)$.

To state bounds we need to define suitable ways of measuring the perturbation of the subspaces, the perturbation of $A$ and the separation of singular values. The usual measure for the distance between two subspaces is the size of the sines of the canonical angles between the subspaces. A comprehensive description of this approach can be found in [1]. If $\Theta$ is the diagonal matrix of canonical angles between two subspaces $\mathscr{W} = \mathscr{R}(W_1)$ and $\widetilde{\mathscr{W}} = \mathscr{R}(\widetilde{W}_1)$ and if

$$\begin{bmatrix} W_1 & W_2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \widetilde{W}_1 & \widetilde{W}_2 \end{bmatrix}$$

are unitary then

$$\|W_2^H \widetilde{W}_1\| = \|\sin \Theta\| = \|W_1^H \widetilde{W}_2\| \tag{9}$$

for any unitarily invariant norm $\|\cdot\|$. The quantities we bound are of the form $\|W_2^H \widetilde{W}_1\|$ so that the resulting theorems can be interpreted as $\sin \Theta$ theorems.

Throughout this paper $\|\cdot\|$ is an arbitrary unitarily invariant norm. If the norm is to be applied to matrices of different sizes, then $\|\cdot\|$ is a family of unitarily invariant norms More explicitly, given a unitarily invariant norm defined for matrices of a given size, the norm is extended to any matrix $X$ of smaller size by

$$\|Xt\| = \left\| \begin{bmatrix} X & 0 \\ 0 & 0 \end{bmatrix} \right\|.$$

For any such family of unitarily invariant norms

$$\|X_{11}\| \leqslant \left\| \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \right\|. \tag{10}$$

The Frobenius norm is $\|\cdot\|_F$ and the spectral norm is $\|\cdot\|_2$. We will regularly use the inequalities

$$\|BC\| \leqslant \|B\|_2 \|C\| \quad \text{and} \quad \|BC\| \geqslant \frac{\|C\|}{\|B^{-1}\|_2},$$

which hold for an arbitrary unitarily invariant norm $\|\cdot\|$.

As one measure of the size of the perturbation of $A$ we use the residuals $R$ and $S$ defined in (3) and a unitarily invariant norm. If the residuals are the result of an additive perturbation $\widetilde{A} = A + E$ then

$$R = -E\widetilde{V}_1 \quad \text{and} \quad S = -E^H \widetilde{U}_1. \tag{11}$$

## 2. The subspace bounds

The proofs of the theorems from [9] depend heavily on projection matrices for singular subspaces. Here we derive related theorems using an approach that more closely parallels the development in [1]. We first represent $U_2^H \widetilde{U}_1$ and $V_2^H \widetilde{V}_1$ as solutions of a matrix equation with the residuals $R$ and $S$ on the right-hand side. A general lemma relating the size of the solution of a matrix equation to the size of the right-hand side then gives the desired bounds.

Multiplying $S = A^H\widetilde{U}_1 - \widetilde{V}_1\widetilde{\Sigma}_1$ by $V_2^H$ on the left and noting that $AV_2 = U_2\Sigma_2$ gives

$$\Sigma_2 U_2^H\widetilde{U}_1 - V_2^H\widetilde{V}_1\widetilde{\Sigma}_1 = V_2^H S. \tag{12}$$

Similarly, multiplying $R = A\widetilde{V}_1 - \widetilde{U}_1\widetilde{\Sigma}_1$ by $U_2^H$ and using $U_2^H A = \Sigma_2 V_2^H$ gives

$$\Sigma_2 V_2^H\widetilde{V}_1 - U_2^H\widetilde{U}_1\widetilde{\Sigma}_1 = U_2^H R. \tag{13}$$

Combining the above gives the matrix equation:

$$\begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix}\begin{bmatrix} V_2^H\widetilde{V}_1 \\ U_2^H\widetilde{U}_1 \end{bmatrix} - \begin{bmatrix} V_2^H\widetilde{V}_1 \\ U_2^H\widetilde{U}_1 \end{bmatrix}\widetilde{\Sigma}_1 = \begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix} \tag{14}$$

relating $V_2^H\widetilde{V}_1$ and $U_2^H\widetilde{U}_1$ to $U_2^H R$ and $V_2^H S$. Multiplying $R$ by $U_3^H$ and noting that $U_3^H A = 0$ completes the set with an equation relating $U_3^H\widetilde{U}_1$ to $U_3^H R$

$$-U_3^H\widetilde{U}_1\widetilde{\Sigma}_1 = U_3^H R. \tag{15}$$

Note that we have already separated out $U_3^H\widetilde{U}_1$ and can conclude that when $\widetilde{\Sigma}_1$ is nonsingular

$$\|U_3^H\widetilde{U}_1\| \leqslant \|\widetilde{\Sigma}_1^{-1}\|_2\|U_3^H R\| \tag{16}$$

for any unitarily invariant norm $\|\cdot\|$. If $\widetilde{\Sigma}_1$ is singular we can draw no conclusion better than the worst case $\|U_3^H\widetilde{U}_1\|_2 \leqslant 1$, i.e. that $\widetilde{U}_1$ contains a vector from the left null space $\mathscr{U}_3$. Since $U_3^H\widetilde{U}_1$ can be bounded by (16), we will from this point on ignore $U_3^H\widetilde{U}_1$.

The right-hand side of (14) involves components $V_2^H S$ and $U_2^H R$ of $R$ and $S$ and not the full residuals. Components of $R$ in the subspaces $\mathscr{U}_1$ and $\mathscr{U}_3$ and components of $S$ in $\mathscr{V}_1$ will not influence the bounds for $U_2^H\widetilde{U}_1$ and $V_2^H\widetilde{V}_1$. The residual $S$ and components of $R$ in $\mathscr{U}_1$ and $\mathscr{U}_2$ do not appear in (16).

Bounding $U_2^H\widetilde{U}_1$ and $V_2^H\widetilde{V}_1$ is simply a matter of applying the technique of [1] to (14). The relevant lemma from [1] is the following.

**Lemma 1.** *If $AX - XB = C$, $\|B\|_2 \leqslant \alpha$ and $\|A^{-1}\|_2 \leqslant (\alpha+\delta)^{-1}$ then*

$$\|X\| \leqslant \frac{\|C\|}{\delta},$$

*where $\|\cdot\|$ denotes any unitarily invariant norm.*

For the Frobenius norm the following stronger result holds [6].

**Lemma 2.** *Suppose $AX - XB = C$ for hermitian matrices $A$ and $B$. If*

$$0 < \delta \leqslant \min_{jk} |\lambda_j(A) - \lambda_k(B)|$$

*then*

$$\|X\|_{\mathrm{F}} \leqslant \frac{\|C\|_{\mathrm{F}}}{\delta}.$$

The proof of the first lemma is the sequence of inequalities

$$\|C\| = \|AX - XB\| \geqslant \|AX\| - \|XB\| \geqslant \|X\|\left(\frac{1}{\|A^{-1}\|_2} - \|B\|_2\right) \geqslant \delta\|X\|.$$

The second lemma follows from the diagonalizations $A = Q_A D_A Q_A^H$ and $B = Q_B D_B Q_B^H$ under which the matrix equation transforms to the diagonal system:

$$[(I \otimes D_A) - (D_B \otimes I)] \operatorname{vec}(Q_A^H X Q_B) = \operatorname{vec}(Q_A^H C Q_B).$$

We start with a perturbation bound using the Frobenius norm. The following theorem follows immediately from (14) and Lemma 2 by noting that the eigenvalues of the hermitian matrix:

$$\begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix}$$

are $\pm\sigma(\Sigma_2)$ so that the singular values, not counting multiplicities, are $\sigma(\Sigma_2)$.

**Theorem 2.** *Let*

$$\delta = \min_{jk} |\sigma_j(\widetilde{\Sigma}_1) - \sigma_k(\Sigma_2)| > 0.$$

*Then*

$$\sqrt{\|V_2^H \widetilde{V}_1\|_F^2 + \|U_2^H \widetilde{U}_1\|_F^2} \leqslant \frac{1}{\delta}\sqrt{\|V_2^H S\|_F^2 + \|U_2^H R\|_F^2}. \tag{17}$$

Except for separate treatment of $U_3^H \widetilde{U}_1$ and the fact that we do not require that $\sigma(\widetilde{\Sigma}_1)$ be separated from zero Theorem 2 is identical to a result from [6].

To obtain results for a general unitarily invariant norm we will use two assumptions on the separation of the singular values of $\widetilde{\Sigma}_1$ and $\Sigma_2$.

1. We assume that there is an interval $[\alpha, \beta]$ and a $\delta > 0$ such that $\sigma(\widetilde{\Sigma}_1)$ is entirely within $[\alpha, \beta]$ and $\sigma(\Sigma_2)$ lies entirely outside of $(\alpha - \delta, \beta + \delta)$. Except for the fact that we do not assume that $\alpha \geqslant \delta$ this is identical to the assumption of Theorem 1.
2. We assume that there is an $\alpha \geqslant 0$ and a $\delta > 0$ such that $\sigma(\widetilde{\Sigma}_1) \leqslant \alpha$ and $\sigma(\Sigma_2) \geqslant \alpha + \delta$. Thus the smallest singular values are in $\widetilde{\Sigma}_1$ and the largest are in $\Sigma_2$. This sorting of the singular values leads to a slightly stronger theorem.

The next two theorems require the first assumption. We assume that the eigenvalues (or equivalently, since $\widetilde{\Sigma}_1$ is hermitian, the singular values) of $\widetilde{\Sigma}_1$ lie in an interval $[\alpha, \beta]$ and use a shift $\widetilde{\Sigma}_1 - \gamma I$ to move the eigenvalues to the interval $[-(\beta - \alpha)/2, (\beta - \alpha)/2]$. This choice of $\gamma$ allows us to apply Lemma 1.

**Theorem 3.** *Suppose that* $0 \leqslant \alpha \leqslant \beta$ *and* $\delta > 0$ *are such that* $\sigma(\widetilde{\Sigma}_1)$ *is contained in* $[\alpha, \beta]$ *and* $\sigma(\Sigma_2)$ *is entirely outside* $(\alpha - \delta, \beta + \delta)$. *Then*

$$\left\| \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \end{bmatrix} \right\| \leqslant \frac{1}{\delta} \left\| \begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix} \right\|, \tag{18}$$

*where* $\|\cdot\|$ *is an arbitrary unitarily invariant norm.*

**Proof.** We consider the shifted equation

$$\left( \begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix} - \gamma I \right) \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \end{bmatrix} - \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \end{bmatrix} (\widetilde{\Sigma}_1 - \gamma I) = \begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix}. \tag{19}$$

If $\gamma = (\alpha+\beta)/2$ then the eigenvalues of $\tilde{\Sigma}_1 - \gamma I$ are in

$$[-(\beta-\alpha)/2, (\beta-\alpha)/2]$$

so that

$$\|\widetilde{\Sigma}_1 - \gamma I\|_2 \leqslant (\beta-\alpha)/2.$$

With this choice of $\gamma$ the eigenvalues of

$$\begin{bmatrix} -\gamma I & \Sigma_2 \\ \Sigma_2 & -\gamma I \end{bmatrix}$$

are outside the interval

$$(-(\beta-\alpha)/2-\delta, (\beta-\alpha)/2+\delta).$$

It follows that

$$\left\| \begin{bmatrix} -\gamma I & \Sigma_2 \\ \Sigma_2 & -\gamma I \end{bmatrix}^{-1} \right\|_2 \leqslant \frac{1}{(\beta-\alpha)/2+\delta}.$$

Lemma 1, with $\alpha$ replaced by $(\beta-\alpha)/2$, applied to (19) gives (18). □

Application of the triangle inequality and simple properties of $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ changes Theorem 3 into something more closely resembling Theorem 1.

**Theorem 4.** *Suppose that $0 \leqslant \alpha \leqslant \beta$ and $\delta > 0$ are such that $\sigma(\widetilde{\Sigma}_1)$ is contained in $[\alpha, \beta]$ and $\sigma(\Sigma_2)$ is entirely outside $(\alpha-\delta, \beta+\delta)$. Then*

$$\max(\|V_2^H\widetilde{V}_1\|, \|U_2^H\widetilde{U}_1\|) \leqslant \frac{2}{\delta}\max(\|V_2^H S\|, \|U_2^H R\|). \tag{20}$$

*If $\|\cdot\|$ is $\|\cdot\|_{\mathrm{F}}$ or $\|\cdot\|_2$ then*

$$\max(\|V_2^H\widetilde{V}_1\|_{2,F}, \|U_2^H\widetilde{U}_1\|_{2,F}) \leqslant \frac{\sqrt{2}}{\delta}\max(\|V_2^H S\|_{2,F}, \|U_2^H R\|_{2,F}). \tag{21}$$

**Proof.** By the triangle inequality

$$\left\|\begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix}\right\| \leqslant \left\|\begin{bmatrix} V_2^H S \\ 0 \end{bmatrix}\right\| + \left\|\begin{bmatrix} 0 \\ U_2^H R \end{bmatrix}\right\| \leqslant 2\max(\|V_2^H S\|, \|U_2^H R\|).$$

By (10)

$$\max(\|V_2^H\widetilde{V}_1\|, \|U_2^H\widetilde{U}_1\|) \leqslant \left\|\begin{bmatrix} V_2^H\widetilde{V}_1 \\ U_2^H\widetilde{U}_1 \end{bmatrix}\right\|.$$

The bound (20) then follows from (18).

To prove (21) we use the general fact that for both $\|\cdot\|_{\mathrm{F}}$ and $\|\cdot\|_2$

$$\left\|\begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix}\right\|_{2,\mathrm{F}}^2 \leqslant \|V_2^H S\|_{2,\mathrm{F}}^2 + \|U_2^H R\|_{2,\mathrm{F}}^2.$$

For $\|\cdot\|_{\mathrm{F}}$ equality clearly holds. For $\|\cdot\|_2$ the inequality follows directly from the definition $\|B\|_2 = \max_{\|x\|_2=1}\|Bx\|_2$. Thus

$$\left\|\begin{bmatrix} V_2^H S \\ U_2^H R \end{bmatrix}\right\|_{2,\mathrm{F}} \leqslant \sqrt{\|V_2^H S\|_{2,\mathrm{F}}^2 + \|U_2^H R\|_{2,\mathrm{F}}^2} \leqslant \sqrt{2}\max(\|V_2^H S\|_{2,\mathrm{F}}, \|U_2^H R\|_{2,\mathrm{F}}).$$

With this improved bound on the residuals, the second inequality (21) follows in the same way as the first. □

Finally, if we separate the singular values into largest and smallest, we can remove the constants 2 and $\sqrt{2}$ in front of the bounds of Theorem 4.

**Theorem 5.** *Suppose that*

$$\sigma_k(\widetilde{\Sigma}_1) \leqslant \alpha \quad \text{and} \quad \sigma_j(\Sigma_2) \geqslant \alpha + \delta \tag{22}$$

*or that*

$$\sigma_k(\widetilde{\Sigma}_1) \geqslant \alpha + \delta \quad \text{and} \quad \sigma_j(\Sigma_2) \leqslant \alpha \tag{23}$$

*for $\delta > 0$ and $\alpha \geqslant 0$ and for all singular values $\sigma_k(\widetilde{\Sigma}_1)$ and $\sigma_j(\Sigma_2)$. Then*

$$\max(\|V_2^H \widetilde{V}_1\|, \|U_2^H \widetilde{U}_1\|) \leqslant \frac{1}{\delta} \max(\|V_2^H S\|, \|U_2^H R\|),$$

*where $\|\cdot\|$ is any unitarily invariant norm.*

**Proof.** We start with the assumption (22) which, combined with (13), gives

$$(\alpha+\delta)\|V_2^H \widetilde{V}_1\| \leqslant \|\Sigma_2 V_2^H \widetilde{V}_1\| \leqslant \|U_2^H \widetilde{U}_1 \widetilde{\Sigma}_1\| + \|U_2^H R\| \leqslant \alpha\|U_2^H \widetilde{U}_1\| + \|U_2^H R\|$$

or

$$(\alpha+\delta)\|V_2^H \widetilde{V}_1\| - \alpha\|U_2^H \widetilde{U}_1\| \leqslant \|U_2^H R\|.$$

From (12) we get

$$(\alpha+\delta)\|U_2^H \widetilde{U}_1\| \leqslant \|\Sigma_2 U_2^H \widetilde{U}_1\| \leqslant \|V_2^H \widetilde{V}_1 \widetilde{\Sigma}_1\| + \|V_2^H S\| \leqslant \alpha\|V_2^H \widetilde{V}_1\| + \|V_2^H S\|$$

so that

$$-\alpha\|V_2^H \widetilde{V}_1\| + (\alpha+\delta)\|U_2^H \widetilde{U}_1\| \leqslant \|V_2^H S\|.$$

Combining these inequalities gives

$$\begin{bmatrix} \alpha+\delta & -\alpha \\ -\alpha & \alpha+\delta \end{bmatrix} \begin{bmatrix} \|V_2^H \widetilde{V}_1\| \\ \|U_2^H \widetilde{U}_1\| \end{bmatrix} \leqslant \begin{bmatrix} \|U_2^H R\| \\ \|V_2^H S\| \end{bmatrix}, \tag{24}$$

where the vector inequality holds componentwise. For $\alpha \geqslant 0$ and $\delta > 0$ the inverse

$$\begin{bmatrix} \alpha+\delta & -\alpha \\ -\alpha & \alpha+\delta \end{bmatrix}^{-1} = \frac{1}{2\alpha\delta+\delta^2} \begin{bmatrix} \alpha+\delta & \alpha \\ \alpha & \alpha+\delta \end{bmatrix}$$

exists and has only positive elements so that we can multiply both sides of (24) by the inverse without changing the inequalities to get

$$\begin{bmatrix} \|V_2^H \widetilde{V}_1\| \\ \|U_2^H \widetilde{U}_1\| \end{bmatrix} \leqslant \frac{1}{2\alpha\delta+\delta^2} \begin{bmatrix} \alpha+\delta & \alpha \\ \alpha & \alpha+\delta \end{bmatrix} \begin{bmatrix} \|U_2^H R\| \\ \|V_2^H S\| \end{bmatrix}.$$

Thus

$$\begin{aligned} \|V_2^H \widetilde{V}_1\| &\leqslant \frac{1}{2\alpha\delta+\delta^2}((\alpha+\delta)\|U_2^H R\| + \alpha\|V_2^H S\|) \\ &\leqslant \frac{1}{2\alpha\delta+\delta^2}(2\alpha+\delta)\max(\|V_2^H S\|, \|U_2^H R\|) \\ &= \frac{1}{\delta}\max(\|V_2^H S\|, \|U_2^H R\|). \end{aligned}$$

Similarly

$$\|U_2^H \widetilde{U}_1\| \leqslant \frac{1}{\delta} \max(\|V_2^H S\|, \|U_2^H R\|).$$

If, instead of (22), we assume (23) then (12) implies

$$(\alpha+\delta)\|V_2^H \widetilde{V}_1\| - \alpha\|U_2^H \widetilde{U}_1\| \leqslant \|V_2^H S\|$$

and (13) implies

$$-\alpha\|V_2^H \widetilde{V}_1\| + (\alpha+\delta)\|U_2^H \widetilde{U}_1\| \leqslant \|U_2^H R\|.$$

Thus the same proof applies with $\|U_2^H R\|$ and $\|V_2^H S\|$ switched. $\square$

These theorems are closely related to the results of [9] and to the Frobenius norm theorem from [6]. Suppose that

$$\hat{\delta} = \min_{jk} \left| \sigma_k(\widetilde{\Sigma}_1) - \sigma_j\left(\begin{bmatrix} \Sigma_2 & 0 \\ 0 & 0 \end{bmatrix}\right) \right|$$

so that $\sigma(\widetilde{\Sigma}_1)$ is separated from the set $\sigma(\Sigma_2) \cup \{0\}$ by $\hat{\delta}$. Then $\hat{\delta} \leqslant \delta$ where $\delta$ is as in Theorem 2 and $\|\widetilde{\Sigma}_1^{-1}\|_2 \leqslant 1/\hat{\delta}$. It follows from Theorem 2 and (16) that

$$\begin{aligned} \|\widetilde{U}_1^H \begin{bmatrix} U_2 & U_3 \end{bmatrix}\|_F^2 + \|\widetilde{V}_1^H V_2\|_F^2 &\leqslant \frac{1}{\hat{\delta}}(\|U_2^H R\|_F^2 + \|U_3^H R\|_F^2 + \|V_2^H S\|_F^2) \\ &\leqslant \frac{1}{\hat{\delta}}(\|R\|_F^2 + \|S\|_F^2), \end{aligned}$$

which is the result from [6]. Separating out $U_3^H \widetilde{U}_1$ does not weaken the bound.

While Theorem 4 does not imply Theorem 1, the same methods used to prove Theorem 4 can be used to prove Theorem 1. Combining the matrix equation (14) with the Eq. (15) for $U_3^H \widetilde{U}_1$ gives

$$\begin{bmatrix} 0 & \Sigma_2 & 0 \\ \Sigma_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \\ U_3^H \widetilde{U}_1 \end{bmatrix} - \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \\ U_3^H \widetilde{U}_1 \end{bmatrix} \widetilde{\Sigma}_1 = \begin{bmatrix} V_2^H S \\ U_2^H R \\ U_3^H S \end{bmatrix}. \tag{25}$$

If $\hat{\delta}$ is chosen so that $\sigma(\widetilde{\Sigma}_1)$ is in $[\alpha, \beta]$ with $0 < \hat{\delta} \leqslant \alpha \leqslant \beta$ and $\sigma(\widetilde{\Sigma}_2)$ is outside $(\alpha - \hat{\delta}, \beta + \hat{\delta})$ then the nearly unmodified proofs of Theorems 3 and 4 applied to (25) gives Theorem 1.

## 3. Singular values and additive perturbations

In the remainder of the paper we assume that $\widetilde{A} = A + E$ and determine the effect of the perturbation $E$ on the singular values in $\widetilde{\Sigma}_1$. In this section we derive first order approximations. We approximate all the singular values contained in $\Sigma_1$ at once and give bounds on the error. In the next section we give a second order approximation to a single perturbed singular value. In all of the results we assume an appropriate separation of singular values and use the results of the previous section which show that $V_2^H \widetilde{V}_1 = O(\|E\|)$, $U_2^H \widetilde{U}_1 = O(\|E\|)$ and $U_3^H \widetilde{U}_1 \widetilde{\Sigma}_1 = O(\|E\|)$.

We need two additional results. The first relates $\|I - QQ^H\|$ to the distance of $Q$ from a unitary matrix.

**Lemma 3.** *Suppose that* $\|I - QQ^H\| = \gamma$ *for some square matrix* $Q$ *and an arbitrary unitarily invariant norm* $\|\cdot\|$. *If* $Q = X\Sigma_Q Y^H$ *is the singular value decomposition of* $Q$ *then*

$$\|XY^H - Q\| \leqslant \gamma.$$

**Proof.** Note that

$$\gamma = \|I - X\Sigma_Q^2 X^H\| = \|I - \Sigma_Q^2\|$$

so that

$$\|XY^H - Q\| = \|X(I - \Sigma_Q)Y^H\| = \|(I + \Sigma_Q)^{-1}(I - \Sigma_Q^2)\| \leqslant \|(I - \Sigma_Q^2)\| = \gamma. \qquad \square$$

The second result is a perturbation bound for singular values [6].

**Theorem 6** (Mirsky). *If*

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \quad \text{and} \quad \tilde{\sigma}_1 \geqslant \tilde{\sigma}_2 \geqslant \cdots \geqslant \tilde{\sigma}_p$$

*are the singular values of two matrices of the same size*, $B$ *and* $\widetilde{B}$, *then*

$$\|\mathrm{diag}(\tilde{\sigma}_i - \sigma_i)\| \leqslant \|\widetilde{B} - B\|$$

*for any unitarily invariant norm* $\|\cdot\|$.

With the singular value decompositions of $A$ and $\widetilde{A}$ defined as in (1) and (2) let

$$U^H(A+E)V = \begin{bmatrix} \Sigma_1 + E_{11} & E_{12} \\ E_{21} & \Sigma_2 + E_{22} \\ E_{31} & E_{32} \end{bmatrix} \tag{26}$$

so that $E_{jk} = U_j^H E V_k$. Then

$$\begin{bmatrix} \widetilde{\Sigma}_1 & 0 \\ 0 & \widetilde{\Sigma}_2 \\ 0 & 0 \end{bmatrix} = \widetilde{U}^H(A+E)\widetilde{V} = \widetilde{U}^H U \begin{bmatrix} \Sigma_1 + E_{11} & E_{12} \\ E_{21} & \Sigma_2 + E_{22} \\ E_{31} & E_{32} \end{bmatrix} V^H \widetilde{V}. \tag{27}$$

Recall that $A$ is $m \times n$ and $\Sigma_1$ and $\Sigma_2$ are $m_1 \times m_1$ and $m_2 \times m_2$ respectively. For a unitary matrix $W$ acting only on the last $m - m_1$ rows of $\widetilde{U}^H U$ we define $S$ by

$$S = W\widetilde{U}^H U = \begin{bmatrix} I_{m_1} & 0 & 0 \\ 0 & W_{22} & W_{23} \\ 0 & W_{32} & W_{33} \end{bmatrix} \begin{bmatrix} \widetilde{U}_1^H U_1 & \widetilde{U}_1^H U_2 & \widetilde{U}_1^H U_3 \\ \widetilde{U}_2^H U_1 & \widetilde{U}_2^H U_2 & \widetilde{U}_2^H U_3 \\ \widetilde{U}_3^H U_1 & \widetilde{U}_3^H U_2 & \widetilde{U}_3^H U_3 \end{bmatrix}.$$

We choose $W$ to introduce zeros in $S$ so that

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & 0 & S_{33} \end{bmatrix} = \begin{bmatrix} \widetilde{U}_1^H U_1 & \widetilde{U}_1^H U_2 & \widetilde{U}_1^H U_3 \\ S_{21} & S_{22} & S_{23} \\ S_{31} & 0 & S_{33} \end{bmatrix},$$

where $S$ is partitioned in the same way as $\widetilde{U}^H U$. Since $\widetilde{U}_2^H U_2$ is square such a $W$ exists and can be found from the $QR$ factorization of the $(m - m_1) \times m_2$ matrix

$$\begin{bmatrix} \widetilde{U}_2^H U_2 \\ \widetilde{U}_3^H U_2 \end{bmatrix}.$$

Applying $S$ instead of $\widetilde{U}^H U$ in (27) gives

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & 0 & S_{33} \end{bmatrix} \begin{bmatrix} \Sigma_1 + E_{11} & E_{12} \\ E_{21} & \Sigma_2 + E_{22} \\ E_{31} & E_{32} \end{bmatrix} \begin{bmatrix} V_1^H \widetilde{V}_1 & V_1^H \widetilde{V}_2 \\ V_2^H \widetilde{V}_1 & V_2^H \widetilde{V}_2 \end{bmatrix} = \begin{bmatrix} \widetilde{\Sigma}_1 & 0 \\ 0 & \widetilde{\Sigma}_{22} \\ 0 & \widetilde{\Sigma}_{32} \end{bmatrix},$$

where we define $\widetilde{\Sigma}_{22}$ and $\widetilde{\Sigma}_{32}$ by

$$\begin{bmatrix} \widetilde{\Sigma}_{22} \\ \widetilde{\Sigma}_{32} \end{bmatrix} = \begin{bmatrix} W_{22} & W_{23} \\ W_{32} & W_{33} \end{bmatrix} \begin{bmatrix} \widetilde{\Sigma}_2 \\ 0 \end{bmatrix}.$$

Applying a permutation to swap the second and third blocks of rows of $S$ and another to swap the second and third blocks of columns we get

$$\begin{bmatrix} S_{11} & S_{13} & S_{12} \\ S_{31} & S_{33} & 0 \\ S_{21} & S_{23} & S_{22} \end{bmatrix} \begin{bmatrix} \Sigma_1 + E_{11} & E_{12} \\ E_{31} & E_{32} \\ E_{21} & \Sigma_2 + E_{22} \end{bmatrix} \begin{bmatrix} V_1^H \widetilde{V}_1 & V_1^H \widetilde{V}_2 \\ V_2^H \widetilde{V}_1 & V_2^H \widetilde{V}_2 \end{bmatrix} = \begin{bmatrix} \widetilde{\Sigma}_1 & 0 \\ 0 & \widetilde{\Sigma}_{32} \\ 0 & \widetilde{\Sigma}_{22} \end{bmatrix}.$$

Since by the construction of $S$, $S_{12} = \widetilde{U}_1^H U_2$ this implies that

$$\begin{bmatrix} \widetilde{\Sigma}_1 \\ 0 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix} \begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix} V_1^H \widetilde{V}_1 + \begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix} \begin{bmatrix} E_{12} \\ E_{32} \end{bmatrix} V_2^H \widetilde{V}_1$$
$$+ \begin{bmatrix} \widetilde{U}_1^H U_2 \\ 0 \end{bmatrix} (E_{21} V_1^H \widetilde{V}_1 + (\Sigma_2 + E_{22}) V_2^H \widetilde{V}_1). \tag{28}$$

Recall that the theorems of the last section guarantee that the matrices $\widetilde{U}_1^H U_2$ and $V_2^H \widetilde{V}_1$ are $O(\|E\|)$ whenever the singular values of $\widetilde{\Sigma}_1$ and $\Sigma_2$ are appropriately separated. If $S_{12} = \widetilde{U}_1^H U_2$ and $V_2^H \widetilde{V}_1$ are $O(\|E\|)$, the matrices

$$\begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix} \quad \text{and} \quad V_1^H \widetilde{V}_1$$

are within $O(\|E\|^2)$ of unitary matrices. Multiplying

$$\widehat{\Sigma} = \begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix}$$

by matrices that are nearly unitary does not greatly change the singular values of $\widehat{\Sigma}$. This can be quantified using multiplicative perturbation theory [3]. However, since this approach would still leave the other additive perturbations, we instead use Lemma 3 and Mirsky's theorem. Except for the first term, everything on the right-hand side is $O(\|E\|^2)$. These terms also have a negligible effect on the singular values.

These observations are made precise, with a bound on the error, in the following theorem.

**Theorem 7.** *Let $A$ and $\widetilde{A} = A + E$ have singular value decompositions* (1) *and* (2) *and let $E_{jk} = U_j^H E V_k$ for $j = 1, 2, 3$ and $k = 1, 2$. Let*

$$\hat{\sigma}_1 \geqslant \hat{\sigma}_2 \geqslant \cdots \geqslant \hat{\sigma}_{m_1}$$

*be the singular values of*

$$\widehat{\Sigma} = \begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix}.$$

*If we assume that* $0 \leqslant \alpha \leqslant \beta$ *and* $\delta > 0$ *are such that* $\sigma(\widetilde{\Sigma}_1)$ *is contained in* $[\alpha, \beta]$ *and* $\sigma(\Sigma_2)$ *is entirely outside* $(\alpha - \delta, \beta + \delta)$ *then*

$$\|\mathrm{diag}(\tilde{\sigma}_i - \hat{\sigma}_i)\| \leqslant \frac{6}{\delta^2}\|E\|_2^2\|A+E\| + \frac{2\sqrt{2}}{\delta}\|E\|_2\|E\|. \tag{29}$$

*If, instead, we assume that*

$$\delta = \min|\sigma(\widetilde{\Sigma}_1) - \sigma(\Sigma_2)|$$

*then*

$$\|\mathrm{diag}(\tilde{\sigma}_i - \hat{\sigma}_i)\| \leqslant \frac{4}{\delta^2}\|E\|_{\mathrm{F}}^2\|A+E\| + \frac{2}{\delta}\|E\|_{\mathrm{F}}\|E\|. \tag{30}$$

*In both cases* $\|\cdot\|$ *is an arbitrary unitarily invariant norm.*

**Proof.** The proof combines the subspace bounds of Section 2 with (28), Lemma 3 and Mirsky's theorem. Since $S$ is unitary

$$\begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix}\begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix}^H = I - \begin{bmatrix} S_{12} \\ 0 \end{bmatrix}\begin{bmatrix} S_{12}^H & 0 \end{bmatrix}$$

with $S_{12} = \widetilde{U}_1^H U_2$. Since $V^H\widetilde{V}$ is unitary

$$V_1^H\widetilde{V}_1\widetilde{V}_1^H V_1 = I - V_1^H\widetilde{V}_2\widetilde{V}_2^H V_1.$$

Thus Lemma 3 implies that there exist $F$ and $G$ with

$$\|F\|_2 \leqslant \|U_2^H\widetilde{U}_1\|_2^2 \quad \text{and} \quad \|G\|_2 \leqslant \|V_1^H\widetilde{V}_2\|_2^2 = \|V_2^H\widetilde{V}_1\|_2^2$$

such that

$$Q = \begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix} - F$$

and

$$P = V_1^H\widetilde{V}_1 - G$$

are unitary. In terms of $Q$ and $P$ (28) becomes

$$\begin{bmatrix} \widetilde{\Sigma}_1 \\ 0 \end{bmatrix} = Q\begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix} P + H, \tag{31}$$

where

$$\begin{aligned} H = {} & F\begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix} V_1^H\widetilde{V}_1 + Q\begin{bmatrix} \Sigma_1 + E_{11} \\ E_{31} \end{bmatrix} G + \begin{bmatrix} S_{11} & S_{13} \\ S_{31} & S_{33} \end{bmatrix}\begin{bmatrix} E_{12} \\ E_{32} \end{bmatrix} V_2^H\widetilde{V}_1 \\ & + \begin{bmatrix} \widetilde{U}_1^H U_2 \\ 0 \end{bmatrix}(E_{21}V_1^H\widetilde{V}_1 + (\Sigma_2 + E_{22})V_2^H\widetilde{V}_1). \end{aligned}$$

Using the bounds on $F$ and $G$ gives

$$\begin{aligned} \|H\| & \leqslant \|U_2^H\widetilde{U}_1\|_2^2\|A+E\| + \|V_2^H\widetilde{V}_1\|_2^2\|A+E\| + \|E\|\|V_2^H\widetilde{V}_1\|_2 \\ & \quad + \|U_2^H\widetilde{U}_1\|_2(\|E\| + \|A+E\|\|V_2^H\widetilde{V}_1\|_2) \qquad (32) \\ & \leqslant (\|U_2^H\widetilde{U}_1\|_2 + \|V_2^H\widetilde{V}_1\|_2)^2\|A+E\| + (\|U_2^H\widetilde{U}_1\|_2 + \|V_2^H\widetilde{V}_1\|_2)\|E\|. \qquad (33) \end{aligned}$$

We have used the fact that (10) implies that

$$\left\|\begin{bmatrix}\Sigma_1+E_{11}\\ E_{31}\end{bmatrix}\right\|\leqslant\|A+E\|$$

as well as the inequalities $\|\widetilde{V}_1^H V_1\|_2\leqslant 1$, $\|E_{21}\|\leqslant\|E\|$ and $\|\Sigma_2+E_{22}\|\leqslant\|A+E\|$.

If the assumptions of Theorem 4 hold then from that theorem and (11) we get

$$\max(\|V_2^H\widetilde{V}_1\|_2,\|U_2^H\widetilde{U}_1\|_2)\leqslant\frac{\sqrt{2}}{\delta}\|E\|_2$$

so that (32) implies that

$$\|H\|\leqslant\frac{6}{\delta^2}\|E\|_2^2\|A+E\|+\frac{2\sqrt{2}}{\delta}\|E\|_2\|E\|.$$

Since $Q$ and $P$ are unitary, (29) follows from an application of Mirsky's theorem to (31).

If the assumptions of Theorem 2 hold then

$$\|V_2^H\widetilde{V}_1\|_{\mathrm{F}}^2+\|U_2^H\widetilde{U}_1\|_{\mathrm{F}}^2\leqslant\frac{2}{\delta^2}\|E\|_{\mathrm{F}}^2,$$

which implies that

$$(\|V_2^H\widetilde{V}_1\|_2+\|U_2^H\widetilde{U}_1\|_2)^2\leqslant 2(\|V_2^H\widetilde{V}_1\|_{\mathrm{F}}^2+\|U_2^H\widetilde{U}_1\|_{\mathrm{F}}^2)\leqslant\frac{4}{\delta^2}\|E\|_{\mathrm{F}}^2.$$

Thus (33) becomes

$$\|H\|\leqslant\frac{4}{\delta^2}\|E\|_{\mathrm{F}}^2\|A+E\|+\frac{2}{\delta}\|E\|_{\mathrm{F}}\|E\|.$$

Another application of Mirsky's theorem establishes (30). □

The bounds show that the singular values of $A+E$ are approximated by those of $\widehat{\Sigma}$ and that the $\mathrm{O}(\|E\|^2)$ error remains $\mathrm{O}(\|E\|^2)$ so long as $\delta$ is not small (i.e. so long as the singular values of $\widetilde{\Sigma}_1$ and $\Sigma_2$ are well separated). By Mirsky's theorem if the singular values in $\Sigma_1$ and $\Sigma_2$ are separated by a distance that is significantly larger than $\mathrm{O}(\|E\|)$ then so are the singular values of $\widetilde{\Sigma}_1$ and $\Sigma_2$.

If $\Sigma_1=\sigma_1$ is $1\times 1$, the theorem gives a new expansion for a perturbed singular value. Since the only singular value of

$$\widehat{\Sigma}=\begin{bmatrix}\sigma_1+u_1^H E v_1\\ U_3^H E v_1\end{bmatrix}$$

is

$$\hat{\sigma}_1=\sqrt{\left|\sigma_1+u_1^H E v_1\right|^2+\|U_3^H E v_1\|_2^2}$$

the theorem states that

$$\tilde{\sigma}_1=\sqrt{|\sigma_1+u_1^H E v_1|^2+\|U_3^H E v_1\|_2^2}+\mathrm{O}(\|E\|^2),\tag{34}$$

where the $\mathrm{O}(\|E\|^2)$ term is bounded by

$$\frac{4}{\delta^2}\|E\|_{\mathrm{F}}^2\|A+E\|+\frac{2}{\delta}\|E\|_{\mathrm{F}}\|E\|_2.$$

We have already noted that a second order approximation to a perturbed singular value does not guarantee accuracy when the singular value is small. Applying the expansion (5) to the example

(7) with $\sigma_1 = 10^{-20}$ and $\epsilon = 10^{-10}$ gives $\tilde{\sigma}_1 \approx 0.5$. In contrast the first order approximation (34) gives

$$\tilde{\sigma}_1 \approx \sqrt{10^{-40} + 10^{-20}},$$

which is the exact perturbed singular value.

## 4. A second order approximation to a perturbed singular value

In this section we derive a second order approximation that is analogous to (34) in that the neglected terms remain $\mathrm{O}(\|E\|^3)$ regardless of the size of $\sigma_1$. The derivation is based on an iteration that generates approximations

$$x_{k+1} = \tilde{v}_1 + \mathrm{O}(\|E\|^{k+1}), \quad y_{k+1} = \tilde{u}_1\tilde{\sigma}_1 + \mathrm{O}(\|E\|^{k+1}) \text{ and } \gamma_{k+1} = \tilde{\sigma}_1 + \mathrm{O}(\|E\|^{k+1})$$

with $\mathrm{O}(\|E\|^{k+1})$ error from approximations $x_k$, $y_k$ and $\gamma_k$ with $\mathrm{O}(\|E\|^k)$ error. The fact that $y_k$ approximates $\tilde{u}_1\tilde{\sigma}_1$ and not $\tilde{u}_1$ is significant. We have already observed that (15) implies that $U_3^H\tilde{u}_1\tilde{\sigma}_1 = \mathrm{O}(\|E\|)$. This is not true of $U_3^H\tilde{u}_1$ which is in general $\mathrm{O}(\|E\|/\tilde{\sigma}_1)$.

In principle the method could be used to generate a formula that approximates $\tilde{\sigma}_1$ with error $\mathrm{O}(\|E\|^k)$ for any $k \geqslant 1$. However even for $k = 3$ the resulting approximation will require some effort to simplify.

We start by specializing (14) to the case in which $\widetilde{\Sigma}_1 = \tilde{\sigma}_1$ is $1 \times 1$ so that the matrix equation becomes

$$\begin{bmatrix} -\tilde{\sigma}_1 I & \Sigma_2 \\ \Sigma_2 & -\tilde{\sigma}_1 I \end{bmatrix} \begin{bmatrix} V_2^H\tilde{v}_1 \\ U_2^H\tilde{u}_1 \end{bmatrix} = \begin{bmatrix} -V_2^H E^H\tilde{u}_1 \\ -U_2^H E\tilde{v}_1 \end{bmatrix},$$

which can be explicitly inverted to get

$$\begin{bmatrix} V_2^H\tilde{v}_1 \\ U_2^H\tilde{u}_1 \end{bmatrix} = \begin{bmatrix} (\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1}(\tilde{\sigma}_1 V_2^H E^H\tilde{u}_1 + \Sigma_2 U_2^H E\tilde{v}_1) \\ (\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1}(\Sigma_2 V_2^H E^H\tilde{u}_1 + \tilde{\sigma}_1 U_2^H E\tilde{v}_1) \end{bmatrix}. \tag{35}$$

As in the rest of this paper we assume that $\tilde{\sigma}_1$ is not in $\sigma(\Sigma_2)$ so that $\tilde{\sigma}_1^2 I - \Sigma_2^2$ is invertible. Notice that it follows from both (35) and from the theorems in Section 2 that $V_2^H\tilde{v}_1 = \mathrm{O}(\|E\|)$ and $U_2^H\tilde{u}_1 = \mathrm{O}(\|E\|)$. We have already noted that $U_3^H\tilde{u}_1\tilde{\sigma}_1 = \mathrm{O}(\|E\|)$.

We are also interested in the quantities $v_1^H\tilde{v}_1$ and $u_1^H\tilde{u}_1$. We assume without loss of generality that $v_1^H\tilde{v}_1$ is real and nonnegative so that

$$v_1^H\tilde{v}_1 = \sqrt{1 - \|V_2^H\tilde{v}_1\|^2} = 1 + \mathrm{O}(\|V_2^H\tilde{v}_1\|^2) = 1 + \mathrm{O}(\|E\|^2).$$

Multiplying the expression for the residual $A\tilde{v}_1 - \tilde{u}_1\tilde{\sigma}_1 = R = -E\tilde{v}_1$ by $u_1^H$ on the left gives

$$\sigma_1 v_1^H\tilde{v}_1 - \tilde{\sigma}_1 u_1^H\tilde{u}_1 = -u_1^H E\tilde{v}_1. \tag{36}$$

We will derive second order approximations $x_2$ and $\gamma_2$ to $\tilde{v}_1 = x_2 + \mathrm{O}(\|E\|^3)$ and $\tilde{\sigma}_1 = \gamma_2 + \mathrm{O}(\|E\|^3)$ by an iteration that starts with the order zero (or higher) approximations

$$x_0 = v_1, \quad y_0 = u_1\gamma_1 \quad \text{and} \quad \gamma_0 = \gamma_1 = \sqrt{|\sigma_1 + u_1^H E v_1|^2 + \|U_3^H E v_1\|^2}.$$

The order zero approximation $\gamma_0$ to the singular value is taken to be the same as the first order approximation $\gamma_1$. To see that $\tilde{v}_1 = x_0 + \mathrm{O}(\|E\|)$ we use the fact that $V$ is unitary to get

$$\begin{aligned}\tilde{v}_1 &= v_1 v_1^H \tilde{v}_1 + V_2 V_2^H \tilde{v}_1 \\ &= v_1\sqrt{1 - \|V_2^H \tilde{v}_1\|^2} + V_2 V_2^H \tilde{v}_1 \\ &= v_1 + \mathrm{O}(\|V_2^H \tilde{v}_1\|) \qquad (37)\\ &= v_1 + \mathrm{O}(\|E\|). \qquad (38)\end{aligned}$$

Since $U$ is also unitary

$$\begin{aligned}\tilde{u}_1\tilde{\sigma}_1 &= u_1 u_1^H \tilde{u}_1\tilde{\sigma}_1 + U_2 U_2^H \tilde{u}_1\tilde{\sigma}_1 + U_3 U_3^H \tilde{u}_1\tilde{\sigma}_1 \\ &= u_1(\sigma_1 v_1^H \tilde{v}_1 + u_1^H E\tilde{v}_1) + \mathrm{O}(\|E\|) \\ &= u_1\sigma_1 + O(\|E\|) = u_1\gamma_1 + \mathrm{O}(\|E\|) \qquad (39)\end{aligned}$$

It will be convenient to represent $x_k$ in the basis provided by $V$ and $y_k$ in the basis provided by $U$ so that the initial approximations are

$$V^H x_0 = \begin{bmatrix}1\\0\end{bmatrix} \quad \text{and} \quad U^H y_0 = \begin{bmatrix}\gamma_1\\0\\0\end{bmatrix}.$$

The iteration that we will use to generate higher order approximations is nominally an iteration for the computation of $\tilde{v}_1, \tilde{u}_1\tilde{\sigma}_1$ and $\tilde{\sigma}_1$. However we will use the iteration only to derive an approximation for fixed $k$ that is valid for small $\|E\|$. We will not need to prove that the iteration converges as $k$ increases. The iteration and the associated errors are described in the following lemma.

**Lemma 4.** *Given approximations* $x_k = \tilde{v}_1 + e_k$, $y_k = \tilde{u}_1\tilde{\sigma}_1 + f_k$ *and* $\gamma_k = \tilde{\sigma}_1 + g_k$ *where* $e_k$, $f_k$ *and* $g_k$ *are* $\mathrm{O}(\delta)$ *define*

$$V^H x_{k+1} = \begin{bmatrix}\sqrt{1 - \|V_2^H x_k\|^2} \\ (\gamma_k^2 I - \Sigma_2^2)^{-1}(V_2^H E y_k + \Sigma_2 U_2^H E x_k)\end{bmatrix},$$

$$U^H y_{k+1} = \begin{bmatrix}\sigma_1\sqrt{1 - \|V_2^H x_k\|^2} + u_1^H E x_k \\ (\gamma_k^2 I - \Sigma_2^2)^{-1}(\Sigma_2 V_2^H E^H y_k + \gamma_k^2 U_2^H E x_k) \\ U_3^H E x_k\end{bmatrix}$$

*and*

$$\gamma_{k+1} = \|(A+E)x_{k+1}\|.$$

*Then*

$$x_{k+1} = \tilde{v}_1 + \mathrm{O}(\delta^2) + \mathrm{O}(\|E\|\cdot\delta), \qquad y_{k+1} = \tilde{\sigma}_1\tilde{u}_1 + \mathrm{O}(\delta^2) + \mathrm{O}(\|E\|\cdot\delta)$$

*and*

$$\gamma_{k+1} = \tilde{\sigma}_1 + \mathrm{O}(\delta^2) + \mathrm{O}(\|E\|\cdot\delta).$$

**Proof.** Since $\gamma_k = \tilde{\sigma}_1 + g_k$ we can expand the inverse to get

$$\begin{aligned}(\gamma_k^2 I - \Sigma_2^2)^{-1} &= [I + (2\tilde{\sigma}_1 g_k + g_k^2)(\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1}]^{-1}(\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1} \\ &= [I - (2\tilde{\sigma}_1 g_k + g_k^2)(\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1}](\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1} + \mathrm{O}(|g_k|^2) \\ &= (\tilde{\sigma}_1^2 I - \Sigma_2^2)^{-1} + \mathrm{O}(|g_k|).\end{aligned}$$

Recall that if $\tilde{\sigma}_1$ is separated from the singular values in $\Sigma_2$ then $(\tilde{\sigma}_1^2 I - \Sigma_2)$ is invertible and the above expansion is valid.

The validity of the lemma can be verified for each of the components of $V^H x_{k+1}$ and $U^H y_{k+1}$. We start with the first component of $U^H y_{k+1}$ and get

$$\begin{aligned}
u_1^H y_{k+1} &= \sigma_1 \sqrt{1 - \|V_2^H x_k\|^2} + u_1^H E x_k = \sigma_1 \sqrt{1 - \|V_2^H(\tilde{v}_1 + e_k)\|^2} + u_1^H E(\tilde{v}_1 + e_k) \\
&= \sigma_1 v_1^H \tilde{v}_1 \sqrt{1 - \frac{e_k^H V_2 V_2^H e_k + 2\mathrm{Re}(\tilde{v}_1^H V_2 V_2^H e_k)}{(v_1^H \tilde{v}_1)^2}} + u_1^H E(\tilde{v}_1 + e_k) \\
&= \sigma_1 \tilde{v}_1^H v_1 + u_1^H E \tilde{v}_1 + \mathrm{O}(\|e_k\|^2) + \mathrm{O}(\|e_k\|\|E\|) \\
&= (\tilde{\sigma}_1 u_1^H \tilde{u}_1 - u_1^H E \tilde{v}_1) + u_1^H E \tilde{v}_1 + \mathrm{O}(\|e_k\|^2) + \mathrm{O}(\|e_k\|\|E\|) \\
&= u_1^H \tilde{u}_1 \tilde{\sigma}_1 + \mathrm{O}(\delta^2) + \mathrm{O}(\delta\|E\|),
\end{aligned}$$

where we have used (36) in line 4 and $\tilde{v}_1^H V_2 = \mathrm{O}(\|E\|)$ in line 3. Note that $(v_1^H \tilde{v}_1)^2 + \|V_2^H \tilde{v}_1\|^2 = 1$ so that $(v_1^H \tilde{v}_1)^2 = 1 - \mathrm{O}(\|E\|^2)$. This ensures that division by $v_1^H \tilde{v}_1$ is harmless.

For $U_2^H y_{k+1}$ we get

$$\begin{aligned}
U_2^H y_{k+1} &= [(\tilde{\sigma}_1^2 I - \Sigma_2)^{-1} + \mathrm{O}(|g_k|)](V_2^H E^H(\tilde{u}_1 \tilde{\sigma}_1 + f_k) + (\tilde{\sigma}_1 + g_k)^2 U_2^H E(\tilde{v}_1 + e_k)) \\
&= (\tilde{\sigma}_1^2 I - \Sigma_2)^{-1}(V_2^H E \tilde{u}_1 \tilde{\sigma}_1 + \tilde{\sigma}_1^2 U_2^H E \tilde{v}_1) \\
&\quad + \mathrm{O}(|g_k|\|E\|) + \mathrm{O}(\|f_k\|\|E\|) + \mathrm{O}(\|e_k\|\|E\|) \\
&= U_2^H \tilde{u}_1 \tilde{\sigma}_1 + \mathrm{O}(\delta\|E\|),
\end{aligned}$$

where we have used (35). For the last component we get

$$U_3^H y_{k+1} = U_3^H E x_k = U_3^H E \tilde{v}_1 + U_3^H E e_k = U_3^H \tilde{u}_1 \tilde{\sigma}_1 + \mathrm{O}(\delta\|E\|)$$

by (15). The proof for $V^H x_{k+1}$ is similar.

For the singular value approximation we have

$$\gamma_{k+1} = \|(A + E)x_{k+1}\| = \|(A + E)(\tilde{v}_1 + e_{k+1})\|$$

so that

$$\|(A + E)\tilde{v}_1\| - \|(A + E)e_{k+1}\| \leqslant \gamma_{k+1} \leqslant \|(A + E)\tilde{v}_1\| + \|(A + E)e_{k+1}\|,$$

where $\|(A + E)\tilde{v}_1\| = \tilde{\sigma}_1$ and $e_{k+1} = \mathrm{O}(\delta\|E\|)$. □

In proving the lemma the only approximations that were made involved the expansion of square roots and the inverse and neglecting products of terms that were $\mathrm{O}(\delta\|E\|)$ and $\mathrm{O}(\delta^2)$. None of these approximations become less accurate as $\tilde{\sigma}_1$ approaches zero.

We partition $U^H E V$ as

$$U^H E V = \begin{bmatrix} e_{11} & e_{12}^H \\ e_{21} & E_{22} \\ e_{31} & E_{32} \end{bmatrix}. \tag{40}$$

To construct $\gamma_2$ we start with

$$V^H x_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad U^H y_0 = \begin{bmatrix} \gamma_1 \\ 0 \\ 0 \end{bmatrix}$$

so that $\gamma_{k+1} = \|(A+E)x_{k+1}\|$ becomes

$$\gamma_{k+1} = \left\| \begin{bmatrix} (\sigma_1 + e_{11})v_1^H x_{k+1} + e_{12}^H V_2^H x_{k+1} \\ e_{21} v_1^H x_{k+1} + (\Sigma_2 + E_{22}) V_2^H x_{k+1} \\ e_{31} v_1^H x_{k+1} + E_{32} V_2^H x_{k+1} \end{bmatrix} \right\|.$$

The first application of the lemma gives the first order approximations

$$V^H x_1 = \begin{bmatrix} 1 \\ (\gamma_1^2 I - \Sigma_2^2)^{-1}(\gamma_1 e_{12} + \Sigma_2 e_{21}) \end{bmatrix} = \begin{bmatrix} 1 \\ \hat{h} \end{bmatrix} = V^H \tilde{v}_1 + \mathrm{O}(\|E\|^2)$$

and

$$U^H y_1 = \begin{bmatrix} \sigma_1 + e_{11} \\ (\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 e_{12}\gamma_1 + \gamma_1^2 e_{21}] \\ e_{31} \end{bmatrix} = \begin{bmatrix} \sigma_1 + e_{11} \\ \hat{g} \\ e_{31} \end{bmatrix} = U^H \tilde{u}_1 \tilde{\sigma}_1 + \mathrm{O}(\|E\|^2),$$

where we define

$$\hat{h} = V_2^H x_1 \quad \text{and} \quad \hat{g} = U_2^H y_1.$$

We do not use the lemma to construct a first order approximation to $\tilde{\sigma}_1$ since we have already defined $\gamma_0 = \gamma_1$ to be a first order approximation.

Another application of the lemma gives

$$V^H x_2 = \begin{bmatrix} \sqrt{1 - \|\hat{h}\|^2} \\ (\gamma_1^2 I - \Sigma_2^2)^{-1}[e_{12}(\sigma_1 + e_{11}) + E_{22}^H \hat{g} + E_{32}^H e_{31} + \Sigma_2 e_{21} + \Sigma_2 E_{22}\hat{h}] \end{bmatrix}.$$

In constructing $\gamma_2$ we note that $x_1 = x_2 + \mathrm{O}(\|E\|^2)$ and neglect third order terms to get

$$\begin{aligned}
\gamma_2 &= \left\| \begin{bmatrix} \sigma_1 + e_{11} & e_{12}^H \\ e_{21} & \Sigma_2 + E_{22} \\ e_{31} & E_{32} \end{bmatrix} V^H x_2 \right\| \\
&= \left\| \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} V^H x_2 + \begin{bmatrix} e_{11} & e_{12}^H \\ e_{21} & E_{22} \\ e_{31} & E_{32} \end{bmatrix} V^H x_1 \right\| + \mathrm{O}(\|E\|^3) \\
&= \left\| \begin{bmatrix} \sigma_1\sqrt{1 - \|\hat{h}\|^2} + e_{11} + e_{12}^H \hat{h} \\ e_{21} + E_{22}\hat{h} + \Sigma_2(\gamma_1^2 I - \Sigma_2^2)^{-1}[e_{12}(\sigma_1 + e_{11}) + E_{22}^H \hat{g} + E_{32}^H e_{31} + \Sigma_2(e_{21} + E_{22}\hat{h})] \\ e_{31} + E_{32}\hat{h} \end{bmatrix} \right\| \\
&\quad + \mathrm{O}(\|E\|^3).
\end{aligned}$$

This can be simplified using the identity

$$I + \Sigma_2^2(\gamma_1^2 I - \Sigma_2^2)^{-1} = \gamma_1^2(\gamma_1^2 I - \Sigma_2^2)^{-1},$$

which implies that

$$(e_{21} + E_{22}\hat{h}) + \Sigma_2(\gamma_1^2 I - \Sigma_2^2)^{-1}\Sigma_2(e_{21} + E_{22}\hat{h}) = \gamma_1^2(\gamma_1^2 I - \Sigma_2^2)^{-1}(e_{21} + E_{22}\hat{h})$$

so that

$$\gamma_2 = \left\| \begin{bmatrix} \sigma_1\sqrt{1 - \|\hat{h}\|^2} + e_{11} + e_{12}^H \hat{h} \\ g + (\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 E_{22}^H \hat{g} + \gamma_1^2 E_{22}\hat{h}] \\ e_{31} + E_{32}\hat{h} \end{bmatrix} \right\| + \mathrm{O}(\|E\|^3)$$

with

$$g = (\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 e_{12}(\sigma_1 + e_{11}) + \Sigma_2 E_{32}^H e_{31} + \gamma_1^2 e_{21}].$$

It turns out that the effect of including the term $(\gamma_1^2 I - \Sigma_2^2)^{-1}(\Sigma_2 E_{22}^H \hat{g} + \gamma_1^2 E_{22}\hat{h})$ is $\mathrm{O}(\|E\|^3)$. To show this we introduce a lemma.

**Lemma 5.** *Let*

$$\breve{\sigma} = \left\| \begin{bmatrix} \sigma \\ \epsilon y \end{bmatrix} \right\| \quad \textit{and} \quad \hat{\sigma} = \left\| \begin{bmatrix} \sigma \\ \epsilon y + \sigma\epsilon^2 x \end{bmatrix} \right\|,$$

*where $\sigma \geqslant 0$ and $\epsilon \geqslant 0$ are real scalars and $x$ and $y$ are vectors. Then*

$$|\hat{\sigma} - \breve{\sigma}| \leqslant \breve{\sigma}\|x\|^2\epsilon^4 + 2\epsilon^3|\mathrm{Re}(y^H x)|.$$

**Proof.** Clearly the claim is true if $\breve{\sigma} = 0$. We assume that $\breve{\sigma} \neq 0$ to get

$$\hat{\sigma} = \sqrt{\sigma^2 + \epsilon^2\|y\|^2 + \breve{\sigma}^2\epsilon^4\|x\|^2 + 2\breve{\sigma}\epsilon^3\,\mathrm{Re}(y^H x)} = \breve{\sigma}\sqrt{1 + \epsilon^4\|x\|^2 + 2\epsilon^3\,\mathrm{Re}(y^H x)/\breve{\sigma}}.$$

The general inequality

$$1 - |\delta| \leqslant \sqrt{1+\delta} \leqslant 1 + |\delta|$$

for real $\delta$, $|\delta| \leqslant 1$ then implies

$$\breve{\sigma}(1 - \epsilon^4\|x\|^2 - 2\epsilon^3|\mathrm{Re}(y^H x)|/\breve{\sigma}) \leqslant \hat{\sigma} \leqslant \breve{\sigma}(1 + \epsilon^4\|x\|^2 + 2\epsilon^3|\mathrm{Re}(y^H x)|/\breve{\sigma}). \qquad \square$$

The point of the lemma is to show that the effect of $\breve{\sigma}\epsilon^2 x$ is $\mathrm{O}(\epsilon^3)$ without expanding the square root

$$\sqrt{1 + \epsilon^4\|x\|^2 + 2\epsilon^3\,\mathrm{Re}(y^H x)/\breve{\sigma}}.$$

It is not apparent that such an expansion would be accurate when $\breve{\sigma}$ is small.

To apply the lemma we let

$$\sigma = \left\| \begin{bmatrix} \sigma_1\sqrt{1 - \|\hat{h}\|^2} + e_{11} + e_{12}^H\hat{h} \\ e_{31} + E_{32}\hat{h} \end{bmatrix} \right\|,$$

$$y = \frac{g}{\|E\|},$$

$$x = (\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 E_{22}^H(\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 e_{12} + \gamma_1 e_{21}] + \gamma_1 E_{22}\hat{h}]/\|E\|^2$$

and $\epsilon = \|E\|$ so that

$$\gamma_2 = \left\| \begin{bmatrix} \sigma \\ \epsilon y + \gamma_1\epsilon^2 x \end{bmatrix} \right\| = \left\| \begin{bmatrix} \sigma \\ \epsilon y + \breve{\sigma}\epsilon^2 x + (\gamma_1 - \breve{\sigma})\epsilon^2 x \end{bmatrix} \right\|,$$

where $\sigma = \sqrt{\sigma^2 + \epsilon^2\|y\|^2}$ is as in the lemma. Since $\gamma_1 = \tilde{\sigma}_1 + \mathrm{O}(\|E\|^2)$ and $\breve{\sigma} = \gamma_2 + \mathrm{O}(\|E\|^2) = \tilde{\sigma}_1 + \mathrm{O}(\|E\|^2)$ we have $(\gamma_1 - \breve{\sigma})\epsilon^2 x = \mathrm{O}(\|E\|^4)$. Thus by the lemma

$$\gamma_2 = \left\| \begin{bmatrix} \sigma \\ \epsilon y + \breve{\sigma}\epsilon^2 x \end{bmatrix} \right\| + \mathrm{O}(\epsilon^4) = \left\| \begin{bmatrix} \sigma \\ \epsilon y \end{bmatrix} \right\| + \mathrm{O}(\epsilon^3),$$

which shows that the effect of $\gamma_1\epsilon^2 x$ on $\gamma_2$ is $\mathrm{O}(\|E\|^3)$. Thus we have proven the following theorem.

**Theorem 8.** *Let $U^H EV$ be partitioned as (40) and let $\gamma_1 = \sqrt{|\sigma_1 + e_{11}|^2 + \|e_{31}\|^2}$. Then*

$$\tilde{\sigma}_1 = \left\| \begin{bmatrix} \sigma_1\sqrt{1-\|h\|^2} + e_{11} + e_{12}^H h \\ g \\ e_{31} + E_{32}h \end{bmatrix} \right\| + \mathrm{O}(\|E\|^3) \tag{41}$$

*with*

$$h = (\gamma_1^2 I - \Sigma_2^2)^{-1}[\gamma_1 e_{12} + \Sigma_2 e_{21}]$$

*and*

$$g = (\gamma_1^2 I - \Sigma_2^2)^{-1}[\Sigma_2 e_{12}(\sigma_1 + e_{11}) + \Sigma_2 E_{32}^H e_{31} + \gamma_1^2 e_{21}].$$

We will now present experimental results comparing various first and second order approximations to a perturbed singular value. The theorems are the simple first order approximation $\tilde{\sigma}_1 = \sigma_1 + e_{11} + \mathrm{O}(\|E\|^2)$, (5) rewritten in terms of the partitioned $E$

$$\begin{aligned}\tilde{\sigma}_1 =& \sigma_1 + e_{11} + \frac{1}{2}\sigma_1 e_{12}^{\mathrm{T}}(\sigma_1^2 I - \Sigma_2^2)^{-1} e_{12} + e_{12}^{\mathrm{T}}(\sigma_1^2 I - \Sigma_2^2)^{-1}\Sigma_2 e_{21} \\ &+ \frac{1}{2}\sigma_1 e_{21}^{\mathrm{T}}(\sigma_1^2 I - \Sigma_2^2)^{-1} e_{21} + \frac{\|e_{31}\|^2}{2\sigma_1} + \mathrm{O}(\|E\|^3)\end{aligned}$$

the result from [5] rewritten in terms of the partitioned $E$

$$\begin{aligned}\tilde{\sigma}_1^2 =& (\sigma_1 + e_{11})^2 + \|e_{21}\|^2 + \|e_{31}\|^2 \\ &+ (\sigma_1 e_{12} + \Sigma_2 e_{21})^H(\sigma_1^2 I - \Sigma_2^2)^{-1}(\sigma_1 e_{12} + \Sigma_2 e_{21}) + \mathrm{O}(\|E\|^3),\end{aligned}$$

the first order approximation

$$\tilde{\sigma}_1 = \sqrt{|\sigma_1 + e_{11}|^2 + \|e_{31}\|^2} + \mathrm{O}(\|E\|^2)$$

from Section 3 and the second order approximation (41).

**Example 2.** We start with an example of the perturbation of a singular value that is not small and for which all approximations give the expected accuracy. Let

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1.7 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} -9.7500\mathrm{e}{-05} & 1.9679\mathrm{e}{-03} & -1.3145\mathrm{e}{-03} \\ 5.6710\mathrm{e}{-04} & -2.1516\mathrm{e}{-03} & -5.1770\mathrm{e}{-04} \\ -4.3790\mathrm{e}{-04} & 1.9823\mathrm{e}{-03} & -1.8330\mathrm{e}{-04} \\ 5.2160\mathrm{e}{-04} & -3.1440\mathrm{e}{-04} & 5.5880\mathrm{e}{-04} \end{bmatrix}.$$

The perturbation was randomly generated and its norm is $\|E\|_2 = 3.6587\mathrm{e}{-03}$. The computation of the perturbed singular value $\tilde{\sigma}_1$ of $A + E$ was done by calling the LAPACK routine DGESVD from a Fortran program compiled with the Intel Fortran compiler on an Intel P4 processor with machine precision $\epsilon \approx 1.1102\mathrm{e}{-16}$. The method is backward stable and computes singular values of a matrix $A + E + F$ where $\|F\|$ is not much larger than $\epsilon\|A\|$. It follows from Mirsky's theorem that the errors in the computed $\tilde{\sigma}_1$ are not much larger than $\epsilon\|A\|$. An estimate $\hat{\sigma}_1$ of the perturbed singular value was computed using each of the five different approximations. Since $\sigma_1^2 I - \Sigma_2$ is well conditioned for this problem, it can be easily verified that the absolute numerical error in computing the various approximations is not much larger than $\epsilon$.

Table 1
Errors in the singular value approximations

| Approximation | $\|\hat{\sigma}_1 - \tilde{\sigma}_1\|$ for Example 1 | $\|\hat{\sigma}_1 - \tilde{\sigma}_1\|$ for Example 2 |
|---|---|---|
| Simple first order | 2.3375e−06 | 6.2734e−04 |
| New first order | 2.4735e−06 | 3.9329e−07 |
| Second order from [5] | 4.6282e−09 | 3.9498e−07 |
| Second order from [7] | 4.8584e−09 | 1.3541e−01 |
| New second order | 5.0385e−09 | 1.3505e−10 |

The unperturbed singular value is $\sigma_1 = 1$. The errors in approximating $\tilde{\sigma}_1$ are summarized in the first column of Table 1. As the approximation from [5] is for $\tilde{\sigma}_1^2$ we have taken the square root of the approximation to get $\hat{\sigma}_1$.

The next column shows the results of applying the theorems to a small singular value. For this example we let

$$A = \begin{bmatrix} 1\text{e}{-06} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1.7 \\ 0 & 0 & 0 \end{bmatrix}$$

and leave $E$ unchanged. The results are summarized in the second column. For each error computed, the effect of numerical error in the computation is $\mathrm{O}(\epsilon)$. In each case, the numerical errors are several orders of magnitude smaller than the quantities in the table. Since we are concerned with comparing the order of magnitude of the errors in each approximation, the numerical errors are not significant.

Notice that the new first and second order approximations retain accuracy as $\sigma_1$ is decreased. We observed in Section 1 that a second order expansion for $\tilde{\sigma}_1^2$ from [5] does not imply a second order approximation to $\tilde{\sigma}_1$ and that a second order expansion to $\tilde{\sigma}_1$ can lose accuracy when $\tilde{\sigma}_1$ is decreased. These facts are apparent in the results for the theorems from [5,7] respectively.

## 5. Multiplicative perturbations

We conclude with a subspace bound for a general multiplicative perturbation of the form:

$$\widetilde{A} = D_1^H A D_2 = (I + F)^H A(I + E) = A + F^H A + AE + F^H AE$$

and an approximation to a singular value in the case of a one-sided multiplicative perturbation

$$\widetilde{A} = AD_2 = A(I + E),$$

where $D_1 = I + F$ and $D_2 = I + E$ are close to the identity. A general multiplicative perturbation can be viewed as an additive perturbation with the special form $F^H A + AE + F^H AE$. The special structure of the perturbation often leads to bounds that are stronger than the corresponding bounds for general additive perturbations. More precisely, it is possible to derive singular subspace bounds that depend on a relative gap between singular values [4] and singular value bounds that bound the relative change in a singular value [3]. Such theorems apply naturally to bound errors in the computation of singular values using a one-sided Jacobi algorithm, which can be shown under certain assumptions to compute the singular values of a matrix $\widetilde{A} = A(I + E)$ where $E$ is small [2].

The approach of using subspace bounds to derive accurate approximations to perturbed singular values can be adapted to the case of a multiplicative perturbation. We start with the subspace bound. In modifying the theorems of Wedin, we have removed the effect of the left null space $U_3$ on the bounds for $U_2^H U_1$ and $V_2^H \widetilde{V}_1$. The same thing can easily be done for the bounds in [4]. The only difference from the derivation in Section 2 is that we must borrow from [4] a lemma and some observations on the special form of the residuals $R$ and $S$ in the right-hand side of (14) In particular $R = A\widetilde{V}_1 - \widetilde{U}_1\widetilde{\Sigma}_1 = (A - \widetilde{A})\widetilde{V}_1$ so that

$$\begin{aligned} U_2^H R &= U_2^H [A(I - D_2) + (D_1^{-H} - I)D_1^H A D_2]\widetilde{V}_1 \\ &= \Sigma_2 V_2^H (I - D_2)\widetilde{V}_1 + U_2^H (D_1^{-H} - I)\widetilde{U}_1\widetilde{\Sigma}_1. \end{aligned}$$

Similarly $S = A^H\widetilde{U}_1 - \widetilde{V}_1\widetilde{\Sigma}_1 = (A^H - \widetilde{A}^H)\widetilde{U}_1$ so that

$$\begin{aligned} V_2^H S &= V_2^H [A^H (I - D_1) + (D_2^{-H} - I)D_2^H A^H D_1]\widetilde{U}_1 \\ &= \Sigma_2 U_2^H (I - D_1)\widetilde{U}_1 + V_2^H (D_2^{-H} - I)\widetilde{V}_1\widetilde{\Sigma}_1. \end{aligned}$$

Thus (14) becomes

$$\begin{aligned} &\begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix} \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \end{bmatrix} - \begin{bmatrix} V_2^H \widetilde{V}_1 \\ U_2^H \widetilde{U}_1 \end{bmatrix} \widetilde{\Sigma}_1 \\ &\quad = \begin{bmatrix} 0 & \Sigma_2 \\ \Sigma_2 & 0 \end{bmatrix} \begin{bmatrix} V_2^H (I - D_2)\widetilde{V}_1 \\ U_2^H (I - D_1)\widetilde{U}_1 \end{bmatrix} + \begin{bmatrix} V_2^H (D_2^{-H} - I)\widetilde{V}_1 \\ U_2^H (D_1^{-H} - I)\widetilde{U}_1 \end{bmatrix} \widetilde{\Sigma}_1. \end{aligned} \tag{42}$$

Bounding $\|V_2^H \widetilde{V}_1\|_F$ and $\|U_2^H \widetilde{U}_1\|_F$ is simply a matter of applying the following lemma from [4] to (42).

**Lemma 6.** *Suppose $AX - XB = AE + FB$ for hermitian matrices $A$ and $B$. If*

$$0 < \delta_2 \leqslant \min_{jk} \left| \frac{\lambda_j(A) - \lambda_k(B)}{\sqrt{|\lambda_j(A)|^2 + |\lambda_k(B)|^2}} \right|$$

*Then*

$$\|X\|_F \leqslant \frac{\sqrt{\|E\|_F^2 + \|F\|_F^2}}{\delta_2}.$$

The result of applying the lemma is the following bound.

$$\begin{aligned} &\sqrt{\|V_2^H \widetilde{V}_1\|_F^2 + \|U_2^H \widetilde{U}_1\|_F^2} \\ &\leqslant \frac{1}{\delta_2}\sqrt{\|U_2^H (I - D_1)\widetilde{U}_1\|_F^2 + \|V_2^H (I - D_2)\widetilde{V}_1\|_F^2 + \|U_2^H (D_1^{-H} - I)\widetilde{U}_1\|_F^2 + \|V_2^H (D_2^{-H} - I)\widetilde{V}_1\|_F^2}, \end{aligned} \tag{43}$$

where

$$\delta_2 = \min_{jk} \left| \frac{\sigma_j(\widetilde{\Sigma}_1) - \sigma_k(\Sigma_2)}{\sqrt{\sigma_j^2(\widetilde{\Sigma}_1) + \sigma_k^2(\Sigma_2)}} \right|.$$

This is simply a result from [4] modified as Wedin's theorems were modified in Section 2 to remove the effect of the left null space. The difference in the resulting theorems is that in the definition of $\delta_2$ in [4], $\sigma_k(\Sigma_2)$ is allowed to take on values in $\sigma(\Sigma_2) \cup \{0\}$, resulting in a potentially smaller value of $\delta_2$. A variety of other results from [4] that apply in the case of an arbitrary unitarily invariant norm can be modified in a similar manner.

We now consider the effect of a one-sided multiplicative perturbation $\widetilde{A} = A(I + E)$ on a singular value $\sigma_1$. Thus $A$ has SVD given by (1) and $\widetilde{A}$ has SVD (2) where $\Sigma_1 = \sigma_1$ and $\widetilde{\Sigma}_1 = \tilde{\sigma}_1$ are $1 \times 1$. It follows that

$$\begin{bmatrix} \tilde{\sigma}_1 & 0 \\ 0 & \widetilde{\Sigma}_2 \\ 0 & 0 \end{bmatrix} = \widetilde{U}^H U \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} (I + V^H E V) V^H \widetilde{V}.$$

If we partition $V^H E V$ as

$$V^H E V = \begin{bmatrix} e_{11} & e_{12}^H \\ e_{21} & E_{22} \end{bmatrix} \tag{44}$$

then

$$\tilde{\sigma}_1 = \begin{bmatrix} \tilde{u}_1^H u_1 & \tilde{u}_1^H U_2 & \tilde{u}_1^H U_3 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 + e_{11} & e_{12}^H \\ e_{21} & I + E_{22} \end{bmatrix} \begin{bmatrix} v_1^H \tilde{v}_1 \\ V_2^H \tilde{v}_1 \end{bmatrix}$$

or

$$\begin{aligned} \tilde{\sigma}_1 =& \sigma_1(1 + e_{11}) \tilde{u}_1^H u_1 v_1^H \tilde{v}_1 + \sigma_1 \tilde{u}_1^H u_1 e_{12}^H V_2^H \tilde{v}_1 \\ & + \tilde{u}_1^H U_2 [\Sigma_2 V_2^H \tilde{v}_1 + \Sigma_2 E_{22} V_2^H \tilde{v}_1 + \Sigma_2 e_{21} v_1^H \tilde{v}_1]. \end{aligned}$$

Since $\widetilde{A} = A(I + E)$ implies $D_1 = I$ and $D_2 = I + E$, (42) gives

$$\begin{aligned} \Sigma_2 V_2^H \tilde{v}_1 =& \tilde{\sigma}_1 U_2^H \tilde{u}_1 - \Sigma_2 V_2^H E \tilde{v}_1 \\ =& \tilde{\sigma}_1 U_2^H \tilde{u}_1 - \Sigma_2 (V_2^H E V)(V^H \tilde{v}_1) \\ =& \tilde{\sigma}_1 U_2^H \tilde{u}_1 - \Sigma_2 \begin{bmatrix} e_{21} & E_{22} \end{bmatrix} \begin{bmatrix} v_1^H \tilde{v}_1 \\ V_2^H \tilde{v}_1 \end{bmatrix}. \end{aligned}$$

Substituting this into the expression for $\tilde{\sigma}_1$ gives

$$\tilde{\sigma}_1 = \sigma_1(1 + e_{11}) \tilde{u}_1^H u_1 v_1^H \tilde{v}_1 + \sigma_1 \tilde{u}_1^H u_1 e_{12}^H V_2^H \tilde{v}_1 + \tilde{\sigma}_1 \tilde{u}_1^H U_2 U_2^H \tilde{u}_1. \tag{45}$$

This relation for $\tilde{\sigma}_1$ is the basis of the following theorem.

**Theorem 9.** *Suppose that the $m \times n$ matrix $A$ with $m \geqslant n$ has rank $n$. Let (1) be the SVD of $A$ and (2) be the SVD of $\widetilde{A} = A(I + E)$ where $\Sigma_1 = \sigma_1$ and $\widetilde{\Sigma}_1 = \tilde{\sigma}_1$ are $1 \times 1$. Define*

$$\delta_2 = \min_k \left| \frac{\tilde{\sigma}_1 - \sigma_k(\Sigma_2)}{\sqrt{\tilde{\sigma}_1^2 + \sigma_k^2(\Sigma_2)}} \right|.$$

*If*

$$\frac{1}{\delta_2^2} \left( \|E\|_2^2 + \|(I + E)^{-H} - I\|_2^2 \right) < 1 \tag{46}$$

*then*

$$\tilde{\sigma}_1 = \sigma_1 \left(1 + e_{11} + \mathrm{O}\left(\frac{1}{\delta_2^2}\|E\|^2\right)\right),$$

*where* $e_{11} = v_1^H E v_1$. *Thus if* $\tilde{\sigma}_1$ *is well separated from* $\sigma(\Sigma_2)$ *in a relative sense then* $\tilde{\sigma}_1 \approx \sigma_1(1 + e_{11})$ *where the neglected terms are second order relative to* $\sigma_1$.

**Proof.** It is easily verified that $\delta_2 \leqslant 1$ so that (46) implies $\|E\|_2 < 1$. The assumption that $A$ has rank $n$ and $\|E\|_2 < 1$ then imply that $\widetilde{A}$ has rank $n$. Thus $\tilde{\sigma}_1 \neq 0$. However (15) gives

$$-U_3^H \tilde{u}_1 \tilde{\sigma}_1 = U_3^H R = U_3^H (A - \widetilde{A})\tilde{v}_1 = U_3^H (A - A(I + E))\tilde{v}_1 = 0$$

so that $\tilde{\sigma}_1 \neq 0$ implies $U_3^H \tilde{u}_1 = 0$. From the fact that $\widetilde{U}^H U$ is unitary we then have

$$|\tilde{u}_1^H u_1|^2 + \|U_2^H \tilde{u}_1\|_2^2 = |\tilde{u}_1^H u_1|^2 + \|U_2^H \tilde{u}_1\|_2^2 + \|U_3^H \tilde{u}_1\|_2^2 = 1$$

so that $|\tilde{u}_1^H u_1|^2 = 1 - \|U_2^H \tilde{u}_1\|_2^2$. By (43) and (46) we have $\tilde{u}_1^H u_1 \neq 0$. Using the bound on $U_2^H \tilde{u}_1$ implied by (43) we have

$$|\tilde{u}_1^H u_1| = 1 + \mathrm{O}\left(\frac{1}{\delta_2^2}\|E\|^2\right) \quad \text{and} \quad |\tilde{u}_1^H u_1|^{-1} = 1 + \mathrm{O}\left(\frac{1}{\delta_2^2}\|E\|^2\right). \tag{47}$$

If we assume that the singular value decompositions are chosen so that $\tilde{v}_1^H v_1$ is real and $\tilde{v}_1^H v_1 > 0$ then we can also show that

$$\tilde{v}_1^H v_1 = 1 + \mathrm{O}\left(\frac{1}{\delta_2^2}\|E\|^2\right). \tag{48}$$

The relation (45) then gives

$$\tilde{\sigma}_1 |\tilde{u}_1^H u_1|^2 = \tilde{\sigma}_1 (1 - \|U_2^H \tilde{u}_1\|_2^2) = \sigma_1 \tilde{u}_1^H u_1 (v_1^H \tilde{v}_1 (1 + e_{11}) + e_{12}^H V_2^H \tilde{v}_1).$$

Thus

$$\tilde{\sigma}_1 |\tilde{u}_1^H u_1| = \sigma_1 |(1 + e_{11}) v_1^H \tilde{v}_1 + e_{12}^H V_2^H \tilde{v}_1| = \sigma_1 \left(1 + e_{11} + \mathrm{O}\left(\frac{1}{\delta_2^2}\|E\|^2\right)\right),$$

where we have used the bound on $V_2^H \tilde{v}_1$ from (43). The theorem then follows from (47) after multiplication by $|\tilde{u}_1^H u_1|^{-1}$. □

The theorem is exactly what might be hoped for on the basis of simpler considerations. Given the perturbation $\widetilde{A} = A + AE$ and the simple first order expansion for a singular value (4) we have

$$\begin{aligned}\tilde{\sigma}_1 &= \sigma_1 + u_1^H A E v_1 + \mathrm{O}(\|E\|^2) = \sigma_1 + \sigma_1 v_1^H E v_1 + \mathrm{O}(\|E\|^2)\\ &= \sigma_1 (1 + v_1^H E v_1) + \mathrm{O}(\|E\|^2).\end{aligned}$$

The significance of the theorem is that it shows that if $\tilde{\sigma}_1$ is well separated from $\sigma(\Sigma_2)$ in a relative sense then the neglected $\mathrm{O}(\|E\|^2)$ terms are small relative to $\sigma_1$.

## References

[1] C. Davis, W. Kahan, The rotation of eigenvectors by a perturbation III, SIAM J. Numer. Anal. 7 (1970) 1–46.
[2] J. Demmel, Applied Numerical Linear Algebra, SIAM, 1997.
[3] I. Ipsen, Relative perturbation results for matrix eigenvalues and singular values, Acta Numerica 1998, Cambridge University Press, Cambridge, 1998, pp. 151–201.
[4] R.-C. Li, Relative pertubation theory. II. Eigenspace and singular subspace variations, SIAM J. Matrix Anal. Appl. 20 (1998) 471–492.
[5] G.W. Stewart, A second order perturbation expansion for small singular values, Linear Algebra Appl. 56 (1984) 231–235.
[6] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, New York, 1990.
[7] J.-G. Sun, A note on simple non-zero singular values, J. Comput. Math. 6 (1988) 258–266.
[8] R.J. Vaccaro, A second-order perturbation expansion for the SVD, SIAM J. Matrix Anal. Appl. 15 (1994) 661–671.
[9] P.-Å. Wedin, Perturbation bounds in connection with singular value decomposition, BIT 12 (1972) 99–111.